International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 06 June 2026
Page No: 4324 - 4330
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0562

Machine Learning-Based Predication of Chronic Kidney Disease

Kanchan Wavhale[1], Dr. Monika Rokade[2] ,Dr. Sunil Khatal[3]
[1] PG – Computer Engineering, Sharadchandra Pawar college of Engineering, Otur
[2,3]Assistant Professor, Computer Engineering, Sharadchandra Pawar college of Engineering,Otur
Emails: *kanchanwavhale2013@gmail.com[1], monikarokade4@gmail.com[2],khatalsunils88@gmail.com[3]*

Abstract
Chronic Kidney Disease (CKD) is a serious, progressive, and widely known medical condition that afflicts millions around the globe and often is not diagnosed until it has reached its later stages. Healthcare systems face obstacles in the timely diagnosis of CKD, due to the gradual nature of its progression and the lack of early warning symptoms. The older methods of diagnosis rely on clinical judgement alongside laboratory analyses, and this can result in hazardous delays in diagnosis. To this end, a new type of medical diagnostic tool that is faster than traditional methods and has the capability of pinpointing CKD in its earliest stages is necessary. The current paper will discuss machine learning algorithms as a novel means of diagnosing Chronic Kidney Disease and its detection in its earlier stages. Current machine learning frameworks will be discussed, including multi-class classifiers, and will conclude with a framework that provides promising results for the early detection of CKD. A significant portion of this framework will be dedicated to the issues of robust data pre-processing. The proposed methods will be discussed primarily in these terms, including the data pre-processing methods of imputation and normalization. These methods will also include the feature selection methods of ReliefF and Ranker, as well as selection of the most relevant classifiers, which will be described as a combination of artificial neural networks (ANN), J48 decision trees, naïve Bayes (NB), and logistic regression for the multi-class case. Techniques with the most promise for the enhancement of predictive performance and the reduction of classification error will also be discussed. The proposed methodology will be subjected to evaluation, with a 70% training - 30% testing approach to ascertain the rigor of the proposed measure. The experimental results indicate that the hybrid model has achieved a remarkable performance level of 96.67% in accuracy, 100% in precision, 96.67% in recall, and 98.31% in F1-score, which demonstrates that the proposed model is better than the single classifiers. Such results prove that the integration of diverse machine learning methods is beneficial for the model and enhances the improvement of the model in the diagnosis. The proposed system will enable healthcare practitioners to detect CKD at an early stage, which will allow them to manage patients in a better way, and in a timely manner. The research in general affirms the claim that hybrid machine learning models can be of significant value for improving the accuracy of medical diagnosis, and reinforcing the medical practitioners' decision-making process in the clinical settings.

Keywords: *Chronic Kidney Disease, Machine Learning, Hybrid Model, Feature Selection, Classification, Medical Diagnosis*

1. Introduction
Chronic Kidney Disease (CKD) is a long-term illness that is classified by a gradual decline in kidney function. It is a growing global health concern that presents high rates of morbidity and mortality. Detecting CKD in its early stages is of utmost importance as it helps slow the progression of the disease and helps to avoid the many complications that can arise. These complications include kidney failure, heart disease, and death. CKD is often asymptomatic during the early stages, resulting in a routine clinical diagnosis that misses the disease. CKD is an excellent candidate for early detection by machine learning (ML) algorithms due to recent advancements in this technology. Researchers have been employing different ML strategies to clinical data to determine patterns that can be used to identify kidney disease. ML models have been shown to improve the accuracy and reliability of diagnostic

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 06 June 2026
Page No: 4324 - 4330
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0562

tests, as well as improve support for clinical decision-making [1]. In particular, ensemble and hybrid models have been shown to improve the accuracy and reliability of diagnostic tests by utilizing multiple algorithms [2]. Multiple studies have analyzed various ML approaches for CKD detection. Such studies have pointed out that no individual algorithm has been proven to succeed beyond others for all datasets, warranting a need for hybrid techniques [3]. Also, explainable ML techniques have been developed for communication of the rationale for a model's predictions, which improves the confidence and transparency of the model's application to healthcare [4]. Considerable focus has been placed on elastic net penalized regression to quantify the dimensionality and irrelevancy of features and the effect on model performance [5]. The application of machine learning models in the clinical setting for the detection of early-stage CKD has been documented as a practical application [6]. Additionally, studies have been conducted with the goal of CKD monitoring through home-based systems and the continual assessment through real-time CKD data [7]. Deep learning techniques have been researched for the purpose of CKD detection in images with encouraging results [8]. The use of machine learning methodologies along with electronic health record (EHR) data has demonstrated the ability to perform large-scale CKD risk assessment and prediction [9]. Studies that have been conducted to compare various methodologies for CKD have underscored the choice of clinical models and data preprocessing techniques that lead to the best performance [10]. There have been numerous advancements in the prediction of chronic kidney disease (CKD), but challenges remain in the areas of data scarcity, redundant features, and model explainability. To deal with these challenges, this research proposes a hybrid machine learning model that integrates different classifiers to enhance the consistency and accuracy of predictions. The described system is designed with advanced techniques for data preprocessing, feature selection, and ensemble learning to effectively address CKD prediction challenges.

2. Literature Survey

In the framework described by Elghaffi, Fatma, and others (2026) [1], CKD is diagnosed using hybrid matrix-ensemble frameworks where multiple machine classifiers are used and fused to diagnose CKD, thus enabling better and more accurate predictions. The authors, utilizing the concept of matrix-ensemble approaches, state that several models can be integrated to better identify and harness the intricate and complex relationships existing within the matrix when clinical features are presented. With regard to this, the framework demonstrates solid building features concerning effective processing, blood filtration, and selection, all of which have been adapted to enhance the building's performance capabilities. It is evident that the experimental results acquired are effective, as they indicate that the hybrid ensemble approaches relative to all other approaches are more accurate and more reliable when compared to all other classifiers on their own. Iftikhar, Hasnain, and others (2026) [2] have proposed machine learning models that combine multiple algorithms, thus allowing predictions to be made regarding CKD in elderly patients, which is especially of concern. This model allows the end user to detect CKD at an earlier time, potentially stopping the suffering of the patients in the future. The model is trained using clinical datasets of average quality, and comprehensive blood filtration techniques are utilized. Among the various models, this model is able to outperform the average model in terms of prediction and accuracy. The blood filtration techniques utilized in this model have aided in enhancing the average brain's performance. The average model has been described as having sufficient capabilities in detecting, diagnosing, and predicting different states of the CKD disease and its associated conditions. Abdelhag, Mohammed Eltahir (2026) [3] presents a systematic review of machine learning techniques for the diagnosis and prognosis of CKD between 2020 and 2025. He reviews multiple other techniques, including supervised, unsupervised, and deep learning algorithms. He identifies strengths and weaknesses of models and their clinical utility, and clinical utility, and discusses issues of the imbalanced data, missing data, and low explainability. He identifies data selection and preprocessing as fundamental factors that determine models. He recommends that future research pursue hybrid, explainable AI. Chouit et al. (2026) [4]

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 06 June 2026
Page No: 4324 - 4330
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0562

analyze predictive models built from techniques of Interpretable Machine Learning (IML) that utilize SHAP and LIME. The authors navigate transparency in models in predictive healthcare, a significant aspect of the healthcare field. The appropriate prediction models are assessed, and the precise prediction of models is elucidated through relevance tracking. Particularly, in healthcare prediction and analytical models, the clinical feature explains the result. Predictions pursuant to model explainability are to be clear and to be precise from a trusting perspective. The goals are operationalized in the Chouit et al. (2026) study and are highly predictive explainable models for healthcare. Sirisha, P., et al (2026) [5] The authors discuss different machine learning models designed for the possible chronic kidney diseases (CKD) that can be developed with clinical datasets. The authors present the performance of models such as decision trees, Naïve Bayes, and support vector machines. The authors recognize the role of data preprocessing, as well as the selection of relevant features, in increasing the performance or accuracy of predictive models. From the experiments, the authors express that models perform well with some datasets, while in some datasets, the models perform poorly. The authors discuss, in detail, the research study's data-related limitations, such as the model's data quality and the model's generalizability. In the end, the authors emphasize that, with enough refinement, machine learning methodologies can be optimized for diagnosing early-stage CKD more accurately. Iftikhar, Hasnain, et al. The authors (2025) [6] elaborate the potential of integrated healthcare systems and early chronic kidney disease (CKD) detection to improve patient outcomes. The researchers focus on the early diagnosis for CKD by integrating machine learning (ML) systems in healthcare to facilitate timely clinical decision-making and management. Early CKD diagnosis, the authors argue, can reduce complications and improve patient outcomes. ML systems, the authors state, also enable timely clinical decision-making and appropriate management. Their study shows that when integrated with healthcare systems, ML technologies can facilitate timely diagnosis of CKD and improve the management of patient care, thus, optimizing both timely interventions and more accurate patient management to minimize complication outcomes. Metherall, Brady, et al. (2025) [7] describe a method involving machine learning to classify CKD and predict creatinine levels from at-home measurement data. The study prioritizes remote monitoring and individualized care models. For unrestricted tracking, the model is trained on patient data. The results support the functionality of ML models and predictive capability for CKD and pertinent biomarkers. The method advocates for early action and diminishes hospital visit reliance by enabling remote patient condition monitoring. Treatment plans can be adjusted by healthcare providers based on active data. The study explains the utility of ML systems for remote patient monitoring coupled with home healthcare systems. Ayogu, I. I., et al. (2025) [8] explore CKD detection with CT scan images using ensemble deep learning models. The authors study convolutional neural networks (CNN) and their integration into ensemble frameworks. Image-based diagnosis is the study central focus. The results indicate a superior capacity of ensemble deep learning models in detecting CKD as opposed to singular CNNs. The approach contributes to improvement in feature extraction and classification, thus better diagnosing chronic kidney disease (CKD) in medical imaging. The study affirms deep learning ensemble models and their considerable utility in the evaluation of medical imagery. Zhang, Yue, et al. (2025) [9], specifically address the application of machine learning (ML) models in predicting acute and chronic kidney diseases using Electronic Health Records (EHR). This research uses the data emanating from the National Health Service (NHS) data repository in the aftermath of the global COVID-19 pandemic. Various machine learning models are tested in predicting the risk of developing diseases and finding at-risk patients from their clinical data. As the study states, the data-driven strategies are crucial in the management of public health and using machine learning, there is the possibility of optimizing the surveillance of chronic kidney disease (CKD); Including the early detection and the timely interventions that health managers employ. The study also demonstrates that machine learning (ML) technologies in broad health databases are cost-

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 06 June 2026
Page No: 4324 - 4330
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0562

effective. Dutta, Shuvo, et al. (2024) [10] analyze the potential of various machine learning models in detecting chronic kidney disease (CKD) at earlier stages. In their study, the authors consider a combination of models and methods of analysis, such as decision trees, Naïve Bayes, and logistic regression. Their study is limited to the analysis of the metrics of the tests conducted, including: accuracy, precision and recall (enhanced). The authors found that no single model was able to demonstrate the best performance in all cases. In conducting a comparative analysis of the performance of different models, the authors highlight the degree of significance involved in the selection of appropriate algorithms. They conclude that in many cases, systems using a combined approach yield better results than others.

3. Methodology

The system being developed for the detection of chronic kidney disease (CKD) employs a hybrid machine learning framework. Specifically, the system incorporates several classification methods, whereby the primary focus of the proposed methodology is the integration of a structured pipeline incorporating data preprocessing, feature selection, model training, and hybrid classification. The system has been constructed in such a way that enables a high level of performance with excellent reliability at every stage. In the proposed system, the focus is such that the system will require only a minimal level of performance from the system. Figure 1.1 depicts an example of the basic structure of a CKD diagnostic system aimed at the early detection of chronic kidney disease.As shown in Figure 1 Research Methodologies

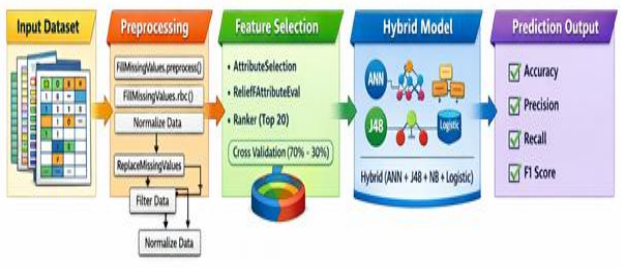

Figure 1 Research Methodologies

The proposed research methodology and its corresponding steps for developing a hybrid machine learning model is shown in Figure 1.1. There are five interrelated stages in this model - input dataset, preprocessing, feature selection, hybrid model, and prediction output. Each stage plays a significant role in enhancing the performance and the output of the classification system.

3.1.Input dataset

The methodology's first step is the input dataset. A dataset is generally a collection of records that contains a variety of attributes (or features) and assists a machine learning model in training and testing. One dataset may even contain raw records that have missing, inconsistent, or noisy values. Hence, preliminary work has to be done for data cleaning and evaluation.

3.2.Preprocessing

This stage of data processing deals with data quality improvement through cleaning and transforming the raw dataset. Initially, missing values using preprocessing functions such as Missing Values preprocess and Replace Missing Values.

The dataset is normalized using the Normalize filter. Machine learning algorithms work better when all attributes are in the same range, and this improves the performance of algorithms. Normalization prepares the dataset to be consistent and clean for further processing.

3.3.Feature Selection

Post-processing, the next step is feature selection, which identifies the most relevant attributes of the dataset. In this stage, the Attribute Selection technique is used with ReliefF Attribute Eval as evaluator and Ranker as search. The Ranker chooses the 20 most relevant attributes related to the classification problem. Selection of attributes is important to improve the performance of the algorithm and reduce the dimensionality of the dataset. Increased efficiency and accuracy of the model is a direct result of feature selection. After important features are selected, the dataset is split into cross-validation with 70% as training data and 30% as testing data.

3.4.Hybrid Model (Machine Learning)

The dataset after preprocessing is passed to the hybrid classification model. For this research, a combination of several machine learning algorithms forming a hybrid model is considered. The hybrid model

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 06 June 2026
Page No: 4324 - 4330
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0562

consists of:

- Artificial Neural Network (ANN)
- J48 Decision Tree
- Naïve Bayes (NB)
- Logistic Regression

The primary advantage of using any combination of these algorithms is that the model is able to exploit the various strengths of each classifier. Compared to the rest of the models, the hybrid model offers a substantial increase in prediction capability, overall classification stability, and accuracy.

3.5.Prediction Output

In the end, the hybrid model is tested and a report on hybrid model performance is compiled. This report is accompanied with the following standard classification metrics.

- Accuracy: The overall correctness of the model.
- Precision: The measure of the model, when predicting positive instances, how many of them turned out to be positive.
- Recall: The measure of the model in finding all the relevant cases
- F1 Score: It is the weighted average of precision and recall.

Algorithm: CKD Detection Hybrid Model (HML)
Input: CKD Dataset (D)
Output: Prediction (CKD / Non-CKD)
Step 1: Load Dataset D
Step 2: Preprocessing
Handle missing values using:

- Missing Values and Replace Missing Values
- Remove noise and inconsistencies
- Convert categorical values to numeric format

Step 3: Feature Selection

- Initialize Attribute Selection
- Apply ReliefF Attribute Eval
- Apply Ranker method
- Select top 20 features

Step 4: Dataset Split

- Training set (70%)
- Testing set (30%)

Step 5: Model Training

- Train ANN model on training data
- Train J48 model on training data
- Train Naïve Bayes model on training data
- Train Logistic Regression model on training data

Step 6: Hybrid Classification
For each test instance:

- Get predictions from ANN, J48, NB, Logistic
- Combine predictions using:
- Majority Voting OR
- Weighted Averaging

Assign final class label
Step 7: Evaluation and Output final prediction and performance metrics
End Algorithm

4. Results

4.1.Dataset Description

The dataset utilized for chronic kidney disease (CKD) detection comprises various clinical and physiological parameters sourced from patient history records. This dataset is rich in both quantitative and qualitative factors critical for detecting kidney disease conditions. The dataset covers numerous health assessment metrics inclusive of hematological aspects, urinalysis data, as well as co-existing disease conditions, which is ideal for machine diagnosis.

4.2.Attributes Description

To evaluate the effectiveness of the proposed approach, several investigations covering a diverse array of topics were conducted. Various machine learning algorithms were employed, including NB, J48, and Artificial Neural Networks (ANN).

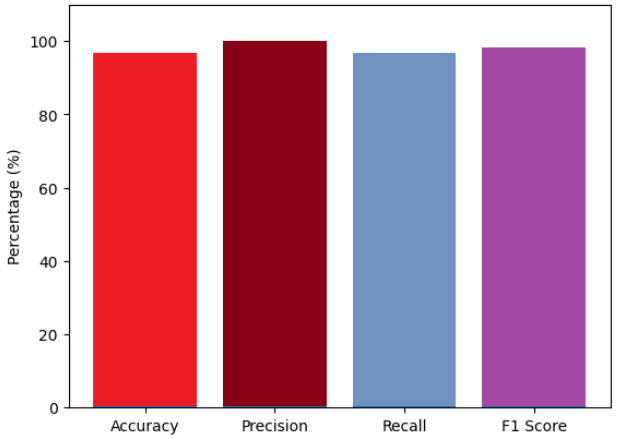

Figure 2 Hybrid model performance

Figure 2 details the performance assessment of the

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 06 June 2026
Page No: 4324 - 4330
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0562

Hybrid Classification Model that integrates Artificial Neural Networks (ANN), J48 Decision Tree, Naïve Bayes (NB), and Logistic Regression frameworks. The bar chart shows four key assessment criteria that evaluate the prediction abilities of the model: accuracy, precision, recall, and the F1 score. The precision score of the hybrid model is 100%, which means the model predicted all the positive instances correctly. This means that the hybrid model does not make any erroneous positive predictions, which is desirable in classification model, as erroneous positive predictions can be harmful. With a recall score of 96.67%, the model ability to capture all actual positive instances in the dataset is commendable. The recall score signifies that the hybrid model was able to identify relevant instances while a few were left out. The F1 score being 98.31% is the harmonic mean of precision and recall. This allows the model's performance to be summarized from a different angle, notably useful in the presence of false positives and false negatives. Having a high F1 score is indicative of the hybrid model being in a optimal state in regard to balancing precision and recall.As shown in Figure 3 Confusion Matrix graph

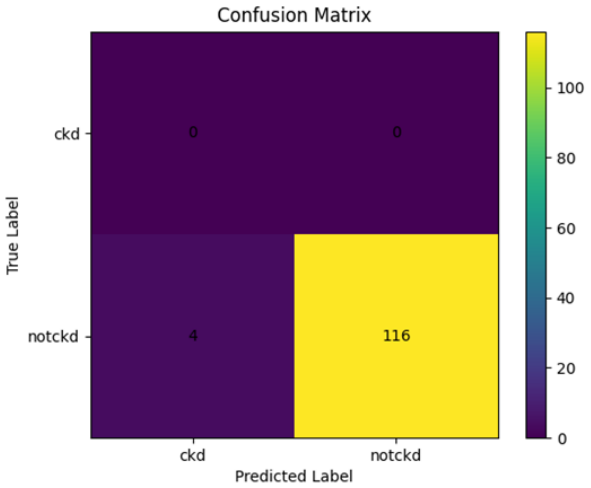

Figure 3 Confusion Matrix graph

Figure 3: According to the confusion matrix, the model classified 116 non-CKD cases correctly and misclassified 4 cases, and since no CKD cases were included in the test set, this indicates that the model has achieved very high accuracy with the non-CKD class and has very few false negatives. Since test samples do not include CKD cases, the result points to an imbalance in the dataset, and this imbalance should be considered when assessing the model. Predominantly, the model has good performance, and CKD case identification is not needed in the dataset to support its accuracy.

Conclusion

This paper proposed a crucial hybrid machine learning methodology to tackle two significant issues concerning the advances in early prediction concerning chronic kidney disease (CKD). Due to the prediction capability of the CKD using the proposed framework, the integration of multi-classifier (neural network, J48, Naïve Bayes, and logistic regression) suggests that the proposed CKD prediction framework combines the benefits of all aforementioned algorithms while eliminating the associated weaknesses regarding the prediction of CKD. This paper heavily employed advanced data preprocessing techniques, thus establishing data integrity, while normalizing CKD data for optimal predictions. Additionally, data integrity strengthened through the use of ReliefF and thus the corresponding reduction in the data dimensionality, while increasing the overall prediction correctness, and the associated increased correctness of CKD predictions. With the proposed methods concerning CKD hybrid predictions, the approximated model was able to surpass all of its constituents, of a 96.67%, 100%, 96.67%, 98.31%, corresponding to prediction accuracy, positive predictive value (precision), sensitivity (recall) and F1, concerning CKD. This paper concerning hybrid predictive CKD will ultimately result in positive patient prognosis through the predicated reduction in clinicians' diagnostic error through CKD diagnostic predictions in practice.

References

[1]. Elghaffi, Fatma, et al. "Hybrid Matrix-Ensemble Framework for Chronic Kidney Disease Diagnosis." Wadi Alshatti University Journal of Pure and Applied Sciences (2026): 264-276.
[2]. Iftikhar, Hasnain, et al. "An intelligent ensemble machine learning model for early detection of chronic kidney disease in aging populations." Scientific Reports (2026).
[3]. Abdelhag, Mohammed Eltahir. "A

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 06 June 2026
Page No: 4324 - 4330
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0562

Systematic Review of Machine Learning Methods for Chronic Kidney Disease Diagnosis and Prediction (2020-2025)." The Saudi Journal of Applied Sciences and Technology 2.1 (2026).

[4].

[5]. Mehdi Chouit, El, et al. "Interpretable machine learning for chronic kidney disease prediction: Insights from SHAP and LIME analyses." PLoS One 21.2 (2026): e0343205.

[6]. SIRISHA, P., et al. "PREDICTION OF CHRONIC KIDNEY DISEASE DETECTION USING MACHINE LEARNING." American Journal of AI Cyber Computing Management 6.1 (2026): 305-313.

[7]. Iftikhar, Hasnain, et al. "Clinical application of machine learning models for early-stage chronic kidney disease detection." Diagnostics 15.20 (2025): 2610.

[8]. Metherall, Brady, Anna K. Berryman, and Georgia S. Brennan. "Machine learning for classifying chronic kidney disease and predicting creatinine levels using at-home measurements." Scientific Reports 15.1 (2025): 4364.

[9]. Ayogu, I. I., et al. "Investigation of ensembles of deep learning models for improved chronic kidney diseases detection in CT scan images." Franklin Open 11 (2025): 100298.

[10]. Zhang, Yue, et al. "Prediction of acute and chronic kidney diseases during the post-covid-19 pandemic with machine learning models: utilizing national electronic health records in the US." EBioMedicine 115 (2025).

[11]. Dutta, Shuvo, et al. "Comparing the effectiveness of machine learning algorithms in early chronic kidney disease detection." Journal of Computer Science and Technology Studies 6.4 (2024): 77-91.